CMIP5 and Multi-Model Ensembles for Climate Research

V. Balaji

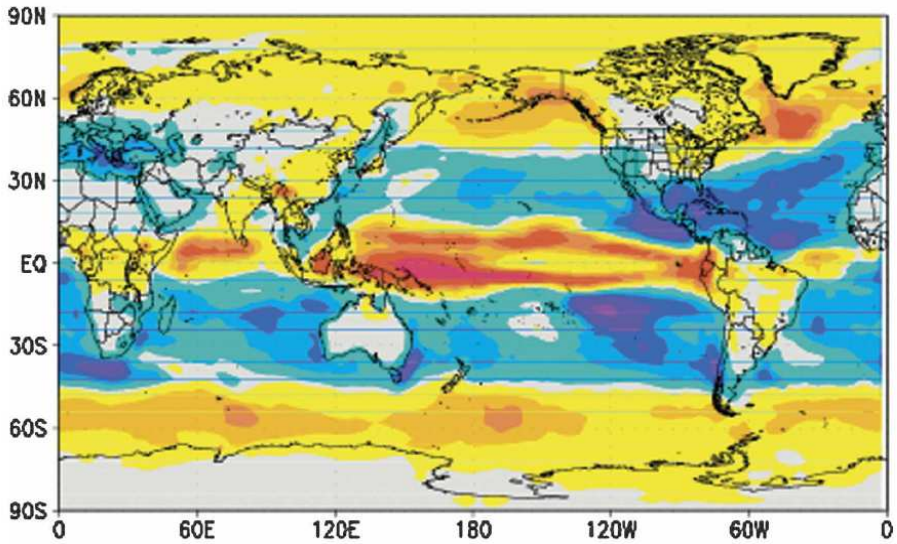
Princeton University and NOAA/GFDL

5 November 2009

The IPCC-AR4 data archive: a global resource

The IPCC data archive at PCMDI is a truly remarkable resource for the comparative study of models. Archives results from $\sim$20 models, used in $\sim$300 papers. . .

Graphics such as this from Held and Soden (2006) are so routinely produced from the IPCC AR4 database that we've ceased to marvel at it. This is a composite of output from 20 models worldwide, run with minimal coordination.

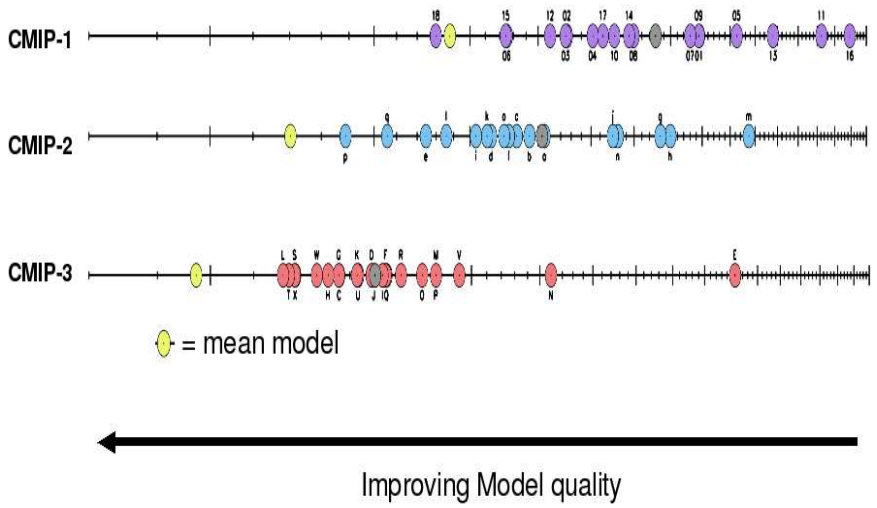


It is worthwhile noting that the ensemble has greater skill by some measure than any individual model (see e.g Reichler et al 2006).

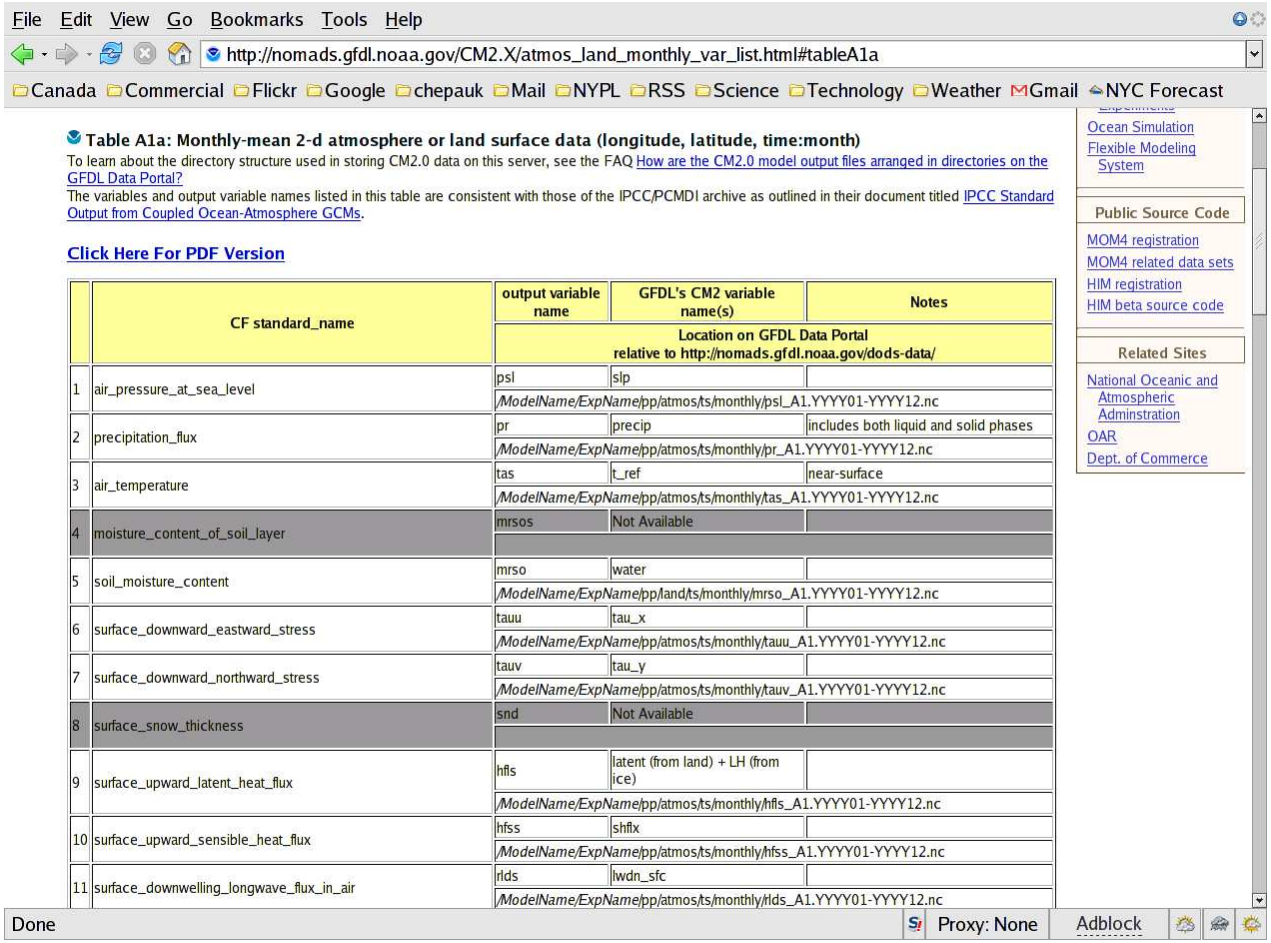
Multi-model ensembles improve predictive understanding

Reichler at al (2006) compare improvement of models' ability to simulate 20th century climate, over 3 generations of models.

- Models are getting better over time.
- The ensemble average is better than any individual model.
- Improvements in understanding percolate quickly across the community.



Data delivery from multi-model ensembles



File Edit View Go Bookmarks Tools Help

http://nomads.gfdl.noaa.gov/CM2.X/atmos_land_monthly_var_list.html#tableA1a

Canada Commercial Flickr Google chepauk Mail NYPL RSS Science Technology Weather Gmail NYC Forecast

Table A1a: Monthly-mean 2-d atmosphere or land surface data (longitude, latitude, time:month)

To learn about the directory structure used in storing CM2.0 data on this server, see the FAQ How are the CM2.0 model output files arranged in directories on the GFDL Data Portal?

The variables and output variable names listed in this table are consistent with those of the IPCC/PCMDI archive as outlined in their document titled IPCC Standard Output from Coupled Ocean-Atmosphere GCMs.

Click Here For PDF Version

	CF standard_name	output variable name	GFDL's CM2 variable name(s)	Notes
		Location on GFDL Data Portal relative to http://nomads.gfdl.noaa.gov/dods-data/		
1	air_pressure_at_sea_level	psl	slp	
		/ModelName/ExpName/pp/atmos/ts/monthly/psl_A1.YYYY01-YYYY12.nc		
2	precipitation_flux	pr	precip	includes both liquid and solid phases
		/ModelName/ExpName/pp/atmos/ts/monthly/pr_A1.YYYY01-YYYY12.nc		
3	air_temperature	tas	t_ref	near-surface
		/ModelName/ExpName/pp/atmos/ts/monthly/tas_A1.YYYY01-YYYY12.nc		
4	moisture_content_of_soil_layer	mrsos	Not Available	
5	soil_moisture_content	mrso	water	
		/ModelName/ExpName/pp/land/ts/monthly/mrso_A1.YYYY01-YYYY12.nc		
6	surface_downward_eastward_stress	tauu	tau_x	
		/ModelName/ExpName/pp/atmos/ts/monthly/tauu_A1.YYYY01-YYYY12.nc		
7	surface_downward_northward_stress	tauv	tau_y	
		/ModelName/ExpName/pp/atmos/ts/monthly/tauv_A1.YYYY01-YYYY12.nc		
8	surface_snow_thickness	snd	Not Available	
9	surface_upward_latent_heat_flux	hfls	latent (from land) + LH (from ice)	
		/ModelName/ExpName/pp/atmos/ts/monthly/hfls_A1.YYYY01-YYYY12.nc		
10	surface_upward_sensible_heat_flux	hfss	shflx	
		/ModelName/ExpName/pp/atmos/ts/monthly/hfss_A1.YYYY01-YYYY12.nc		
11	surface_downwelling_longwave_flux_in_air	rlds	lwdn_sfc	
		/ModelName/ExpName/pp/atmos/ts/monthly/rlds_A1.YYYY01-YYYY12.nc		

Ocean Simulation
Flexible Modeling System

Public Source Code
MOM4 registration
MOM4 related data sets
HIM registration
HIM beta source code

Related Sites
National Oceanic and Atmospheric Adminstration
OAR
Dept. of Commerce

Done Proxy: None Adblock

- increased reliance of federated database and petabyte-scale distributed archives.
- Critically depends on software, metadata, and data standards.

Metadata standards: an unsung hero

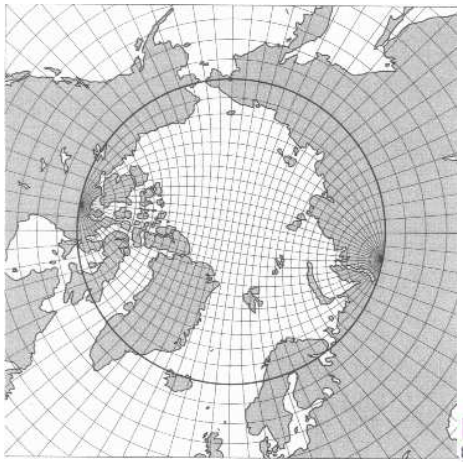
The unglamorous and mostly unfunded activity of building metadata standards proceeds under the guidance of informal grassroots activities, recently acknowledged as central by WMO working groups WGCM and WGNE:

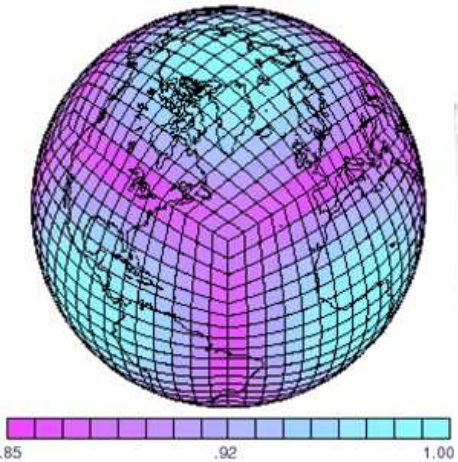
- CF Conventions: `http://cf-pcmdi.llnl.gov/`
- GO-ESSP Consortium: `http://go-essp.gfdl.noaa.gov/`
- METAFOR: `http://metaforclimate.eu`

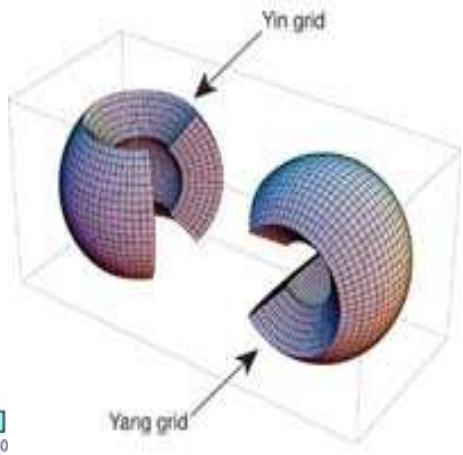
Models and experiments e.g IPCC AR4.

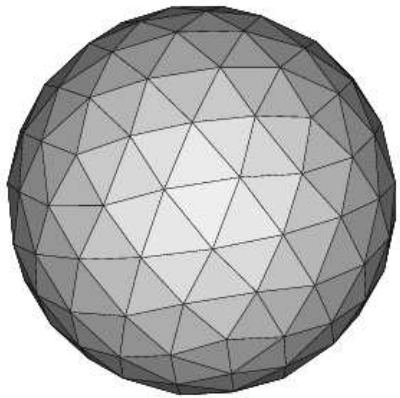
Variable names e.g Temperature with units kelvin.

Model grids time and space and planetary geometry.



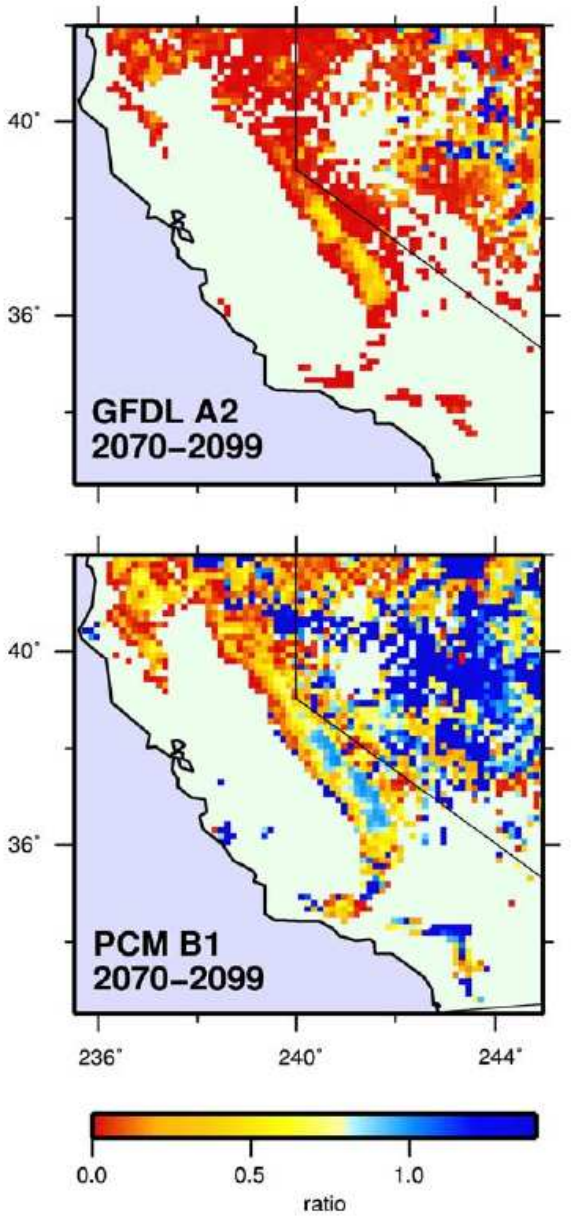






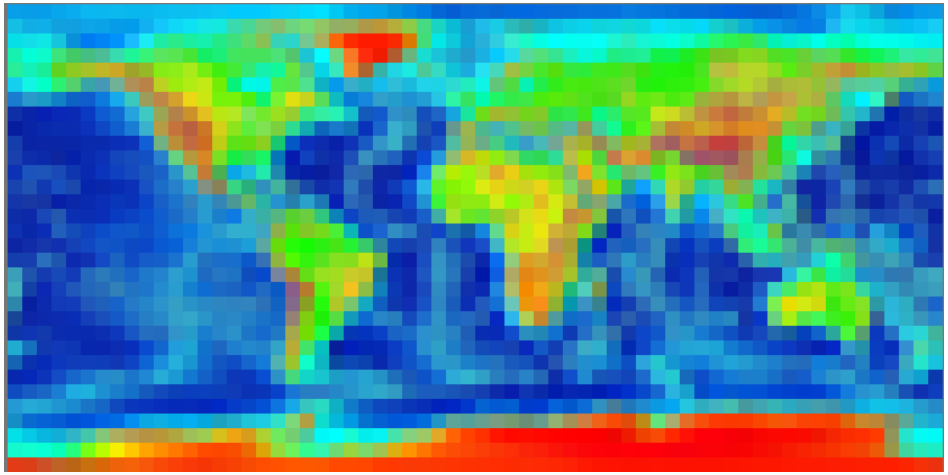
Downstream science: statistical downscaling

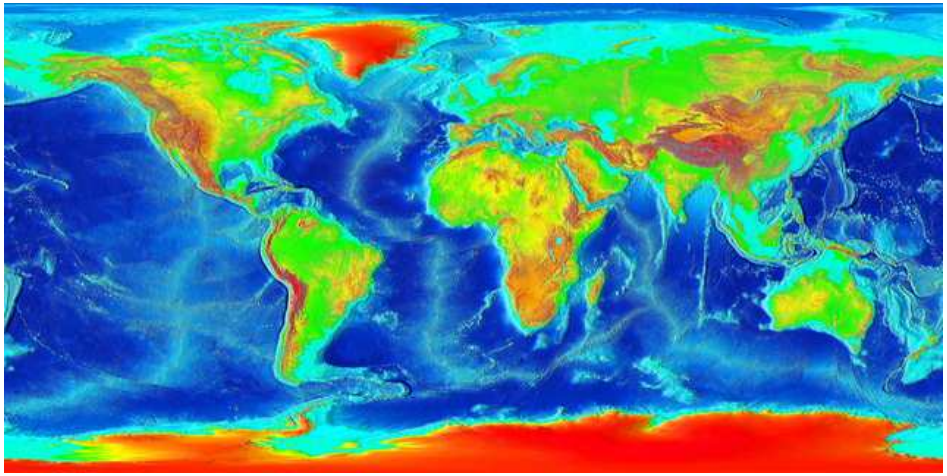
- Cayan et al (2008): downscaled AR4 model output used to drive hydrology model (VIC).
- Required rerunning model at GFDL with modified outputs (dailies saved for 2070-2099 from A2 run).
- Potential service activity if the service has access to computing and archival resources.



Predictive capability is limited by computing power

- For each 10X increase in resolution (number of pixels or *grid cells*), we require roughly a 1000X increase in computing power.
- For a given amount of computing power, we pick a model resolution that affords sufficient throughput (say, 100 years per month).
- Current-generation models:
  - run on 100-1000 processors
  - generate about 10-100 TB per century (1 DVD = 4.7 GB)
  - typical experiment like IPCC-AR4 requires about 5000 model years.





What's new in AR5

- Many modeling centres may run at higher resolution: for instance, our atmospheric time-slice experiments will run at 25 km resolution..
- CMIP5 requests surface data at high temporal resolution (daily, 6h, 3h) for key time periods (late 20th, mid 21st).
- New experiments and variables: e.g carbon chemistry.
- Distributed network of servers with single sign-on: expected volumes too large to be all stored at PCMDI.
- Native grid data.
- Support for "WG2 users" a key new focus: deployment of on-the-fly services for regridding, subsetting, GIS-friendly output, canned analyses, downscaling. . .

Current plan for CMIP5/AR5 data distribution

- Federated data distribution involving 3 gateway nodes (PCMDI, BADC and MPI DDCs) and several data nodes. Core datasets replicated at gateways. PCMDI will still host data upon request by modeling center.
- Projected data volumes: 0.5-1 PB at gateways, smaller fractions at data nodes. Data requests made at ESG nodes will be resolved where the dataset resides.
- Common metadata for model fields and grids processed by ESG Publisher and catalogued by THREDDS. Model metadata (component descriptions, physics options, forcing datasets) captured by CMIP5 Questionnaire developed by METAFOR in structured, searchable form.
- "IPCC Expert Meeting on Assessing and Combining Multi Model Climate Projections" to provide advice on using the whole MME.

GIP climate themes

- Multi-model ensembles and distributed (federated) data archives are a central methodology in the field. The archives are a resource both for domain and downstream scientists. Enabling these archives (resources, standards) is a CI activity (e.g rerunning a model with a different output data profile).
- A framework that enables scientist-produced analysis software (1-100 PE range code, eg “NINO3 SST spectra”, comparisons with obs, downscaling) to be run “near” the data archive would be a transformative new element in CI.

`http://gip.noaa.gov/references/paper_0907_-nesiishortv12.doc`

Thank you! Questions?